NASA TECHNICAL TRANSLATION

NASA TT F-11,252

NASA TT F-11,252

SPEECH COMMANDS IN CONTROL SYSTEMS

E. Kyunnap

Translation of "Ustnyye Komandy v Sistemakh Upravleniya" Izvestiya Akademii Nauk Estonskoy SSR, Seriya Fiziko-Matematicheskikh i Tekhnicheskikh Nauk, Vol. 15, No. 3, pp. 377-399, 1966

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, DC 20546 APRIL 1968

NASA TT F-11,252

SPEECH COMMANDS IN CONTROL SYSTEMS

E. Kyunnap

ABSTRACT. A review of the literature dealing with automatic recognition of speech sounds. The problem of increasing channel carrying capacity is considered. A study is made of the mechanism of sound formation. The principles of operation of band-pass, formant, scanning, harmonic and correlation voice coders are outlined. A number of signal devices for recognizing speech signals are described, and the use of universal computers as a means of studying and recognizing speech signals is discussed.

1. Multiplexing of the Communications Channel

As is known, a speech signal consists of a sum of individual oscillations of various frequencies and amplitudes. When it is expanded into a series, summation may be performed either with respect to elements equally spaced by frequency (Fourier series) or with respect to elements equally spaced by time (Kotel'nikov's theorem). In the first case, the speech signal is subjected to harmonic analysis and the amplitudes (and phases) of the harmonics are transmitted through the channel; at the receiving point, the speech signal is restored using these spectral coefficients, determined by the analyzer at the transmitting end of the communications channel. In the second case, pulses are transmitted through the communications channel at discrete time intervals; the amplitudes of the pulses are proportional to the instantaneous values of the function, read at intervals Δt. At the receiving end, these pulses pass through a filter, the output of which is constantly added. /377[1]

The information transmitted contains, in addition to the useful information, a certain quantity of noise. A criterion characterizing the signal level in comparison to the noise level is the value $H = \log p/p_n$, where p and p_n are the mean powers of signal and noise respectively. The product of three quantities V = TFH is called the signal volume. A communications channel is also characterized by three quantities: T_k, the time interval during which the channel is connected; F_k, the band of frequencies transmitted through the channel; and H_k, the power level of the apparatus making up the channel. The product $V_k = T_k F_k H_k$ is called the channel capacity. The condition $V_k \geqslant V$ must be assured if the signal is to pass through the channel.

[1] Numbers in the margin indicate pagination in the foreign text.

Multiplexing of a communications channel can be achieved by deforming one of the pairs of these quantities while leaving the third quantity unchanged. Deformation of the signal is performed by compressing it at the transmitting /378 end of the channel and expanding it at the receiving end. Changes in H, T and F correspond to changes in amplification, delay of the signal using a delay line and frequency respectively. For example, if a speech signal is recorded on magnetic tape, transmitted at double speed, re-recorded at the receiving end and then played back at half speed, the volume of information transmitted remains unchanged, but transmission is performed twice as fast at twice the frequency.

Companding of a speech signal, i.e. compression at the transmitting end and expansion at the receiving end of a communications channel, may be frequency, amplitude or time companding. One extreme form of amplitude companding of a speech signal is clipping. In this case, the amplitude of the speech signal is limited at two levels and only the points at which the function changes its sign are transmitted.

As we know, the smaller the base of a number system, the greater the number of digits required to represent the same numbers. An optimal system would be a system with the base e. However, it is impossible in practice to produce such a system, and the base used must be either 2 or 3. The most widely used system is the binary system, although the trinary system would be more efficient, since 3 is closer than 2 to the value of e. Speech clipping (Figure 1) corresponds to a vinary number system of transmission. As I. Licklider has stated [111, 112], when a speech signal is limited to two levels, a sufficient quantity of information still remains in the signal to provide intelligibility at the required level. The technical conditions for impulse telephony call for 128 levels. Consequently, clipping a speech signal decreases its volume by a factor of 7.

The intelligibility of speech increases if the speech signal is differentiated before clipping. With this technique, the frequency of the clipped signal is increased and the location of the extreme points of the original signal is transmitted. Partial companding is achieved by dividing the frequency of this signal at the transmitting end and correspondingly multiplying it at the receiving end [36, 156]. The spectrum is narrowed by only a factor of 6, and great distortion of the speech results; therefore, this companding is not particularly promising. However, reduction of the frequency range of speech by limiting the upper and lower ends of the spectrum has the opposite effect. The human voice occupies a frequency range from 50-60 Hz to 15-20 KHz. If the only demand placed on a telephone conversation is intelligibility and recognition of the voice of the person speaking, the upper frequency limit can be reduced to 2.5-3 KHz. If intelligibility alone is required, the signal volume can be decreased still further. Under noise conditions, limiting the frequencies transmitted over a telephone channel to a maximum of 3500 Hz and a minimum of 300 Hz increases intelligibility of the speech [31, 130, 136].

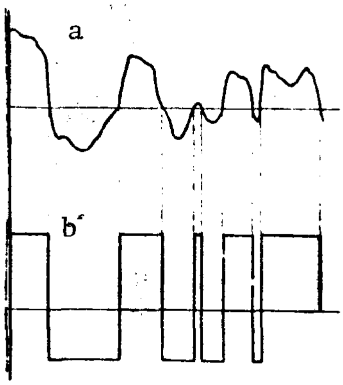

Figure 1. Clipping of a Speeeh Signal: a, Original speech signal; b, Clipped speech signal

Time companding of speech eliminates certain time intervals, and the pauses which arise are filled with other transmitted material. At the receiving end, the pauses in the speech are filled in by the listener mentally [25, 36, 39, 73]. However, the low deg-ee of companding achieved (up to 2) and the reduction in intelligibility indicates that this method has no particular future prospects [103]. Better results for time multiplexing of communications channels are yielded by using the pauses between words and phrases in natu-al speech, as well as the pauses on one line while the conversation partner speaks on the other line. The switching of individual conversations on one line into the pauses in other conversations can be performed by clipping the speech signal. With this system, four to six conversations can be transmitted through one communications channel.

Parametric companding methods, although they allow considerably greater compression of a communications channel, disrupt the microstructure of the speech signal: only the parameters produced by a speech signal analyzer are transmitted through the channel, and at the receiving end these parameters are used to control a speech synthesizer. Thus, the parametric methods of companding involve automatic recognition and synthesis of speech signals. /379

The devices used to transmit the speech signal by the parametric method have come to be called vocoders [59,60]. In semi-vocoder devices, the parametric method is used to transmit only the upper portion of the speech signal, and the lower portion is transmitted continually [69]. In addition to providing multiplexing of communications channels, vocoders can be used to provide secrecy for telephone conversations [2, 102, 157].

2. Investigation of the Formation of Sound

A number of works have been dedicated to the investigation of the functioning of the ear [14, 84, 92, 114] and the vocal cords [64, 89, 94, 165]. In [117], six male voices are investigated by applying Fourier analysis to one oscillating period of pressure in the speech channel, while in [105, 106], the dependence of the base tone on air pressure in the glottic chink is determined, and in [24], sound formation and the determination of formants by anatomical measurements of the vocal tract using X-rays and electronic computer processing of the data are investigated.

Speech is created by pulses from the vocal cords, the oscillating frequency of which determines the pitch of the base tone. These sound pulses have a discrete spectrum with a large number of harmonics over a broad frequency range. The frequency of the base tone lies primarily between 80 and 350 Hz. According to the data of some authors, the amplitudes of the harmonics are almost identical over a broad frequency range [17], while other authors indicate that the amplitude of the harmonics decreases regularly with increasing frequency [64]. The audible signal is produced from the sound of the vocal cords as it passes through the resonating cavities of the mouth and nose, as a result of which the amplitudes of certain harmonics of the vocal cord sound are decreased or completely suppressed, while that of others is reinforced, forming resonant peaks (so-called formants), the measurement of which has also been the subject of a number of works [65, 66, 165].

Each voiced sound corresponds to its own combination of formants. According to the data of [3], in Russian speech the vowels oo, oh, ah and ee are characterized by a single formant, while the sound eh has two and the sound y[1] -- 3. According to the data of [64, 66, 165], good recognition of vowels requires that the first three formants be determined; according to [129], only two need be determined. Unvoiced consonants have no clearly expressed formant areas and are distinguished by the amplitudes of the zero (M_0), first (M_1) and second (M_2) orders, characterizing the spectrum in the band selected [22]:

$$M_0 = \Sigma A_n, \quad M_1 = \Sigma f_n A_n, \quad M_2 = \Sigma f_n^2 A_n^2,$$

where A_n is the amplitude of the n-th band of the spectrum, f_n is its mean frequency.

The sound of the vocal cords h(t) has the frequency spectrum of the even and odd portions of the resonators, i.e.

$$H(j\omega) = \int_{-\infty}^{\infty} h(t) e^{-j\omega t} dt.$$

The resonators of the oral and nasal cavity, depending on the speech sound being formed, have transfer functions $G(j\omega)$, i.e. $G_a(j\omega)$, $G_o(j\omega)$, $G_y(j\omega)$, etc., corresponding to the vowels a, o, y, etc. The signal, upon leaving the mouth, is determined by the equation

$$c(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(j\omega) H(j\omega) e^{j\omega t} d\omega.$$

[1] A vowel sound with no exact equivalent in English -- Tr.

Since the sound of the vocal cords has a spectrum almost identical for all people, the audible signal is determined only by the transfer function of the resonator $(C(j\omega) = G(j\omega)H(j\omega)$. Each phoneme corresponds to its own standard transfer function. On the basis of this phenomenon, a device has been developed for recognizing a man from his voice [188], as well as for producing information concerning the condition of an astronaut during flight [1]. /380

Each sound has its own shades. The timbre of the voice of any man differs depending on the properties of the resonators and the change in the base tone during speech. The amplitudes and frequencies of the formants of each phoneme may change within certain limits [87, 113, 131]. This fact makes more difficult the design of a device for automatic recognition and synthesis, since the same concentrations of energies at the same frequency may belong to different phonemes [110].

There are various opinions concerning the significance of the formants. Some investigators believe that their definition can have no decisive significance in developing a device for recognizing speech signals [79, 121]; however, most authors still hold the opposite opinion [24, 67]. For example, if we record the vowel ah on magnetic tape and suppress certain formant areas within it, the ah is converted, for example, to an oo.

As we know, most phonemes, including the vowels, can be reproduced by a whisper, without using the vocal cords. The exciting action is the breathing noise, which can be looked upon as a random, stationary process $n(t)$, with spectral density $N(\omega)$. Then, the spectral density of the output quantity, i.e. the phoneme produced, for example ah, will be

$$N_a(\omega) = |G_a(j\omega)|^2 N(\omega).$$

As we can see from this formula, the base tone carries no information concerning the phoneme, and therefore in developing an apparatus for automatic speech recognition, there is no need to take the base tone into consideration.

The transfer function of a phoneme $G(j\omega)$ contains both amplitude and phase data. As we know, the hearing does not react to a change in the phase shifts of a complex signal; therefore, in developing an apparatus for automatic recognition, these shifts can be ignored; however, in synthesis, a phase shift between harmonics worsens the quality of speech sound considerably [35, 41, 108, 147].

Dynamic spectrographs (so-called videographs) have been developed for the analysis of speech signals; these videographs make it possible to produce a three-dimensional representation of speech: the abscissa represents time, the ordinate represents frequency and the third coordinate, the darkness of a point on the frequency-time plane, represents the amplitude of the signal at the given frequency [128, 137, illeg.] (Figure 2).

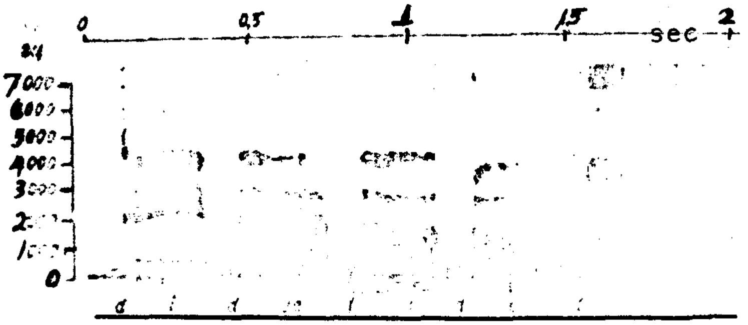

Figure 2. Videograph of Syllables (example taken from [64])

/381 Videographs are divided according to their principle of operation into devices with sequential analysis and devices with parallel analysis. One of the first videographs was developed by H. Sund [166]. This device also makes it possible to photograph the amplitude-frequency dependence on a cathode ray tube. One defect of Sund's videograph is the impossibility of observing and recording the time envelopes of the speech signal.

One modification of the videograph is the intervalograph [44]. In this device, a signal is produced whose amplitude is proportional to the intervals between transitions of the speech signal through the zero level.

On request by the Institute of Language and Literature of the Academy of Sciences ESSR, the former Scientific Research Electronic Engineering Institute of the ESSR Council of the National Economy prepared a spectrograph with parallel analysis. A set of 52 filters is used to cover the range from 40 Hz to 14 KHz. The spectrogram of the speech signal being analyzed is produced simultaneously on three cathode ray tubes, making it possible to photograph and visually observe changes in the spectrum of the speech signal.

A new variant of a spectrograph manufactured by General Electric has 80 filters with a spectral width of 75 Hz each, a local oscillator, a scanning device, an amplitude logarithmizing device, and a camera for photographing the results of analysis. The spectrograph can also be used for analysis of the songs of birds, machine noises, the operation of the heart, etc. [186, 187].

Using the videograph it has been established that the accuracy of recognition is increased not only by determining the position of formants and their intensity, but also by establishing the rate with which they change in frequency and level.

Speech is a continuous function of time between pauses for breathing, and consists of individual, discrete phonetic elements -- phonemes. The phonemes are varieties of sounds which depend on their pronunciation. There are always more phonemes than sounds. In the Russian and English languages there are approximately 40, in Estonian there are approximately 30 and in German there are approximately 40 [47].

The problem of recognition of speech sounds can be solved either by comparison of one phoneme, syllable or word from the corresponding set stored in the memory of the machine, or by acoustical characteristics alone. In the first case, the machine performs comparison of the characteristics of the input speech signal with the characteristics of signals stored in its machine memory, and the input signal is output as the sound with the greatest correlation coefficient. In the second case, the output signal is formed by analysis alone.

The perception of speech by man can be divided into three stages. In the first stage, the acoustical stage, a number of physical phenomena are perceived and a combination of parameters is determined; in the second stage, the phonetic stage, these parameters are compared with standard parameters in memory and the initial recognition of the speech sound is performed; finally, in the third, linguistic stage, the content of the information produced is clarified (for example, endings are added to words which the speaker may not have pronounced at all, etc.).

Automatic recognition of speech sounds is based primarily on the performance of the first two stages. In the general case, the machine must contain devices storing information concerning the language in order for complete recognition of speech sounds to be possible.

The most widespread methods of recognition of speech sounds are analysis of the spectral, time and spectral-time characteristics.

The spectral method was first described by L. Myasnikov [16, 17, 18], then later by others as well [61, 129]. According to his suggestion, the audible oscillations are analyzed by pairs of filters with the following pass-bands: 500-700 and 800-1000 Hz; 1250-1500 and 4000-5000 Hz; 650-750 and 5500-6500 Hz; 1250-1500 and 400-500 Hz. The output of each filter pair is detected and fed in counterphase to an indicator whose arrow either remains at the central position or is deflected to one side or the other. The phonemes are differentiated only according to the frequency, regardless of the amplitude. The replacement of the indicator with the arrow by a three-position, polarized relay was used to produce a recognition accuracy of up to 75-80%.

C. Smith used 32 filters in his device [160, 161], and employed the principle of accentuation of formant peaks by determining the amplitude differences in the adjacent filters. The signals produced were fed to a comparison system which reacted only to signals with maximum amplitude. A certain combination of channel numbers, i.e. formant frequencies, corresponds to a certain vowel phoneme. However, the results were unsatisfactory due to the fact that the uncertainties resulting from the influence of neighboring phonemes on each other and displacement of formants resulting from changes in the base tone could not be eliminated. /382

3. Vocoders

The investigation of formants and the spectral method of analysis were used as the basis for construction of band, formant, scanning, harmonic and correlation vocoders.

In the band type vocoder, the entire range of the speech signal is analyzed in band filters having either even frequency division throughout the entire range or even frequency division only in the lower range (up to 1000 Hz) with logarithmic frequency division in the higher range (above 1000 Hz) [104]. The outputs from each filter are rectified and used as parameters for analysis of the speech signal.

In the first vocoder [59], the speech signal was analyzed by ten main and two supplementary filters. The band width of the first filter was 250 Hz, of the others -- 300 Hz; the entire range analyzed was 2950 Hz wide. The output from each filter was rectified, passed through a supplementary low frequency filter (tuned to 0-250 Hz) and transmitted through a communications channel to the synthesizer. The synthesizer was a noise generator with a frequency limit of 250-3500 Hz. The output of the generator was connected to the generator filter, which had the same frequency ranges as the analyzer. Thus, the output of the generator filter was controlled by corresponding output of the analyzer filter signal. The base tone was separated from the signal before it passed through the analyzer filters and passed through a second, supplementary filter with a frequency pass-band of 0-50 Hz for smoothing.

Since in this vocoder all the higher harmonics of the speech signal are not transmitted, the synthesized speech sounds rigid and hoarse, but the intelligibility is satisfactory.

The communications channel carries ten main signals and two supplementary signals. Each channel requires a band width of approximately 25 Hz, i.e. a total of approximately 300 Hz.

If we compare uncompressed speech signals with identical volumes of information but with different ratios of frequency and dynamic ranges, we have

$$P_{c_2}/P_{n_2} = (P_{c_1}/P_{n_1})^{F_1 c_1/F_2 c_2},$$

whereP_{c_1}, P_{n_1}, P_{c_2} and P_{n_2} are the power of signal and noise at the transmitting and receiving ends of the communications channel; F_1 and F_2 are the corresponding frequency bands; c_1 and c_2 are the throughput capacities for transmission of quantities of information.

Assuming that $c_1 = c_2$, i.e. the throughput capacities of ordinary and vocoder transmission are identical, where $F_1 = 3000$ Hz and $F_2 = 300$ Hz we have

$P_{c_2}/P_{n_2} = (P_{c_1}/P_{n_1})^{10}$, i.e. theoretically the vocoders require ten times greater dynamic range. Actually, this requirement is overstated and, according to the data of a number of authors [35, 63], if transmission of uncompressed speech requires a signal/noise ratio of about 30 db, a band type vocoder with ten channels requires a ratio of about 40 db [161].

Vocoder signals can be transmitted either continuously or in the form of pulses. The continuous signal has a spectral width of not over 15-50 Hz. The dynamic range is somewhat less than that of the original speech, not exceeding 25-30 db. With pulse transmission, vocoder signals are quantized by level and time. In frequency, not over two samples per Hertz are required, while in amplitude, samples each 1-1.5 db will suffice. With a spectral width of 25 Hz and a dynamic range of about 16 db, the channel should provide a capacity of 200 bits per second (25 X 2 = 50 samples; $2^4 = 16$ and 4 X 50 = 200) or a total throughput capacity of about 2000 bits per second for a ten-channel vocoder.

We know that the transmission of pulses requires a frequency of no less than 1-1.5 Hz per bit/sec, so that transmission at a rate of 2000 bit/sec requires a channel with a frequency width of 3000 Hz. /383

According to information theory, speech can be transmitted without distortion when the following condition is fulfilled:

$$A > \frac{1}{3} D_{av} F \quad \text{bit/sec}$$

where A is the throughput capacity of the channel; D_{av} is the average effective dynamic range, F is the width of the frequency range of speech. If we consider that the speech signal range is 5000 Hz, the mean dynamic range is 30 db, then $A \geqslant 5 \cdot 10^3$ (bit/sec); consequently, according to the data above, the loading of communications channels can be decreased by a factor of approximately 25.

In a semi-vocoder, the lower frequency of speech (up to 600 Hz) is transmitted without conversion, while the high frequency area (above 600 Hz) is analyzed and transmitted by a vocoder [69, 155]. In comparison to vocoders, the required channel volume in this case is increased, but the intelligibility of one-syllable words is increased from 74 to 84%.

In a scanning vocoder [175, 176, 178], the speech signal is analyzed in 100 filters, the outputs of which are detected and stored in condensers. The voltage level of the condensers is transmitted by a rotating commutator at 30 rotations per second to a synthesizer, where the sound is restored using a controlled multivibrator. Vocoders have also been developed in which the base tone is not transmitted [34]; in this case, the intelligibility produced is up to 62.5%.

transmitted in linear combinations. The difference is in the synthesis at the receiving end: in band type vocoders, the parameters are fixed by filters with the same bands as the analyzers, while harmonic vocoders have filters at the receiving end corresponding to the expansion into the Fourier series.

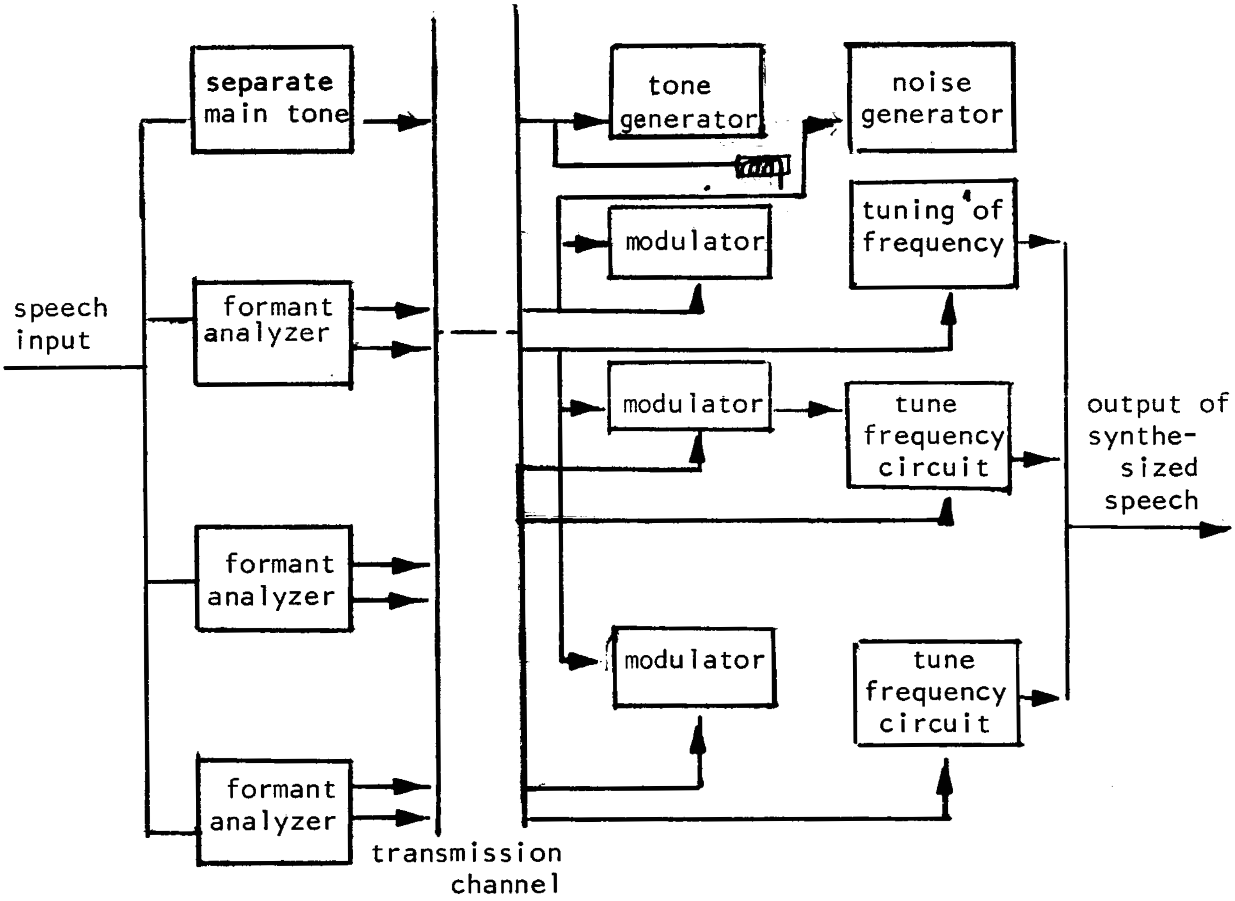

Figure 3. Block Diagram of Formant Vocoder

The correlation method of analysis is based on the relationships between the autocorrelation function R(τ) and the energy spectrum of the signal S(ω) [23, 33, 154]:

The correlation method allows us to avoid the influence of the effects of phased shifts in the synthesized speech, which appear in band type, formant

In a pulse vocoder [175], there are ten evenly distributed filters, the outputs of which are rectified and used to control pulse generators so that pulse width modulation is produced, the number of pulses for each formant being proportional in amplitude and frequency. According to the statement of the author of [175], this vocoder has great interference stability.

The spectral-time method of speech recognition differs from the spectral method in that the output of the filters is scanned with respect to time as well as frequency. As a result, phonemes are transmitted through the communications channel in the form of code symbols [140, 142].

A further development of the spectral method is the formant method. It consists of determination of the presence of a given formant in a given band of filters [67-69, 75] (Figure 3). In vocoders of this type, up to four formants are distinguished. Improved formant vocoders, in which the intensity as well as correlation between frequencies and amplitudes of formants are analyzed, have been developed by several authors [3, 50, 64]. Formant vocoders in which the moments of first and second order are considered have given better results, particularly for determination and synthesis of consonants than the vocoders described above [43, 63, 85]. It has been suggested that the first formant be determined both in the 250-850 Hz range and in the 300-1200 Hz range. According to [85], it is sufficient to have an upper limit to the filter defining the first formant of not over 1000 Hz. The second formant is determined between 900 and 2300 Hz. Since the frequencies of formants overlap in this vocoder, it has been suggested that the position of the first formant be determined first. If it is between 700 and 900 Hz, the frequency of the second formant must be no lower than 1400-1800 Hz; if, however, the first formant is considerably lower than 700 Hz, the frequency of the second formant lies between 700 and 900 Hz and it is necessary to retune the filters accordingly. If a third formant is transmitted as well, it will be defined in the 2100-3500 Hz band, the fourth -- in the 4000-6000 Hz range. /384
The moments are determined by differentiation of the spectrum. If the base tone is transmitted by two parameters (frequency and level) and the four formants are transmitted by three parameters, 14 signals must be transmitted in all. If the moments as well as the dynamic indicators of the formants such as rate and direction of change of frequency and level are transmitted, the number of signals transmitted is increased.

Scanning vocoders also have analyzing filters, but the signals are transmitted sequentially in time to the expander [177]. Whereas in formant vocoders the speech signal is analyzed, in contrast to band vocoders, only at frequencies corresponding to the position of the formants in the speech signal, harmonic vocoders analyze the speech signal completely, determining the Fourier coefficients and transmitting the terms of the series (except for the constant component) through the communications channel using two parameters. At the receiving end, these parameters control either a discrete spectrum generator, or a noise generator, sometimes both[20, 21, 27]. In their principle of operation, harmonic vocoders differ little from band type vocoders. In both cases, the ordinates of the spectrum are determined, except that in band vocoders they are transmitted without conversion, while in harmonic vocoders they are and harmonic vocoders due to the presence of a complex impedance in the band filters of the synthesizer. This achieves compression by a factor of 10 [153, 154]. /385

In order to improve intelligibility of the phonemes in the spectral-time method, it has been suggested that all phonemes be preliminarily divided according to their characteristic indicators [94, 95]. An electronic binary system [43, 45, 185] separates voiced and unvoiced sounds using filters -- the voiced sounds have a base tone, the unvoiced sounds do not. In the next step, the noise voiced sounds are distinguished from the non-noise voiced sounds by the presence or absence of the first formant at the output of the filters. Unvoiced sounds are divided into plosive and fricative by the difference in their amplitudes. The unit for separating voiced sounds into noise and non-noise types operates with an accuracy of up to 95%. The unit for separating vowels into higher and lower vowels operates with an accuracy up to 98%, while the unit separating lower vowels into diffuse and compact operates with an accuracy of up to 94%.

Comparative data on the volume of the communications channels are shown in Table 1 [159].

TABLE 1

Coding method	Required channel volume, bit/sec
discrete form of speech signal	30,000[1]
phoneme	60
word (120 words per minute)	
a) vocabulary of 2 words	2
b) vocabulary of 8,000 words	26
vocoder	2,000
teletype (120 words per minute)	75

[1] Considering that the speech signal range is 3,000 Hz.

A number of works have been dedicated to improvement of the technical indicators of vocoders [49, 114, 164]. For example, digital vocoders have been developed where in place of exciting the synthesizer with the voice, as was done in the first vocoders, the excitation is performed by the base tone which in turn is used to turn on a special generator. The synthesizer uses a reducing device, which eliminates the noise arising as a result of variation of the base tone [171]. A new system of compression has been developed which does not require separation of the base tone [76]. The usage of a multi-channel modulator has reduced the transmitted signal spectrum to 1.4 KHz [77]. A vocoder system has been suggested allowing a considerable reduction in the volume of the communications channel (to 94 bit/sec); the operation of the vocoder is based on the usage of a memory device in which all phonemes are stored, each phoneme having its own code and being "called up" by digital signals [47].

Another suggestion for reducing the volume of the communications channel is based on the usage of syllable synthesizers. A code for the syllables in the form of digits is obtained in the analyzer, and the digits are then transmitted to the synthesizer and the required syllable is selected on the basis of these code digits. A synthesizer with a volume of 200 syllables has been constructed on this basis [126].

The EVA electron-analog synthesizer reproduces speech sounds using curves drawn with current conducting ink on a tape or drum [96]. The transmission of speech by this method can be performed at a frequency 30 times lower than that required for ordinary telephone conversation. The intelligibility of words reaches 75%. According to a statement by the author of [96], intelligibility can be increased to 85%, and the frequency compression increased to 1,000. This type of transmission requires very little power, which is extremely important in establishing communications with astronauts.

4. Special Devices for Recognition of Speech Signals

The spectral-time method was studied in detail by H. Olson and H. Belar [122, 123] and J. Dreyfus-Graf [57, 58], who developed typewriters controlled by oral commands. The first machine of H. Olson and H. Belar typed words from /386
a vocabulary of ten single-syllable words in the memory of the machine; the third typewriter had a vocabulary of 150 single-syllable words (Figure 4). In the opinion of the authors, a typewriter with a memory of 1,000 sound combinations is sufficient for practical uses [124] (obviously, this requirement is also sufficient for the Russian and Estonian languages, since in these languages phonetic transcription and printed form differ little from each other). The latest model of their voice powered typewriter consists of eight band-pass filters, eight amplitude-comparing detectors, syllable and orthographic memory units, control devices and the typewriter. The filters cover the range from 250 to 20,000 Hz. The output of each channel is transmitted to a device where the maxima are accentuated by comparing the levels of the outputs of neighboring filters and the second derivative of the curve of the speech signal is defined. Quantization of the input signal with respect to time occurs each 0.2 sec, and after its output from the filters it is
quantized in the memory each 0.04 sec (in all, also 0.2 sec) and is quantized /387
by amplitude in three levels.

The comparison circuit compares the output curve of the speech signal before and after quantization, i.e. each 0.04 sec. If no change in signal amplitude occurs during this time, the signal does not reach the memory. The memory has 256 different characters, including pause. The results of analysis are transmitted in an eight digit code, which at a rate of pronunciation of ten phonemes per second requires a rate of 80 bits per second. The accuracy of operation of the machine is 92-94% if it is adjusted to the speaker using it, considerably lower if random speakers are used.

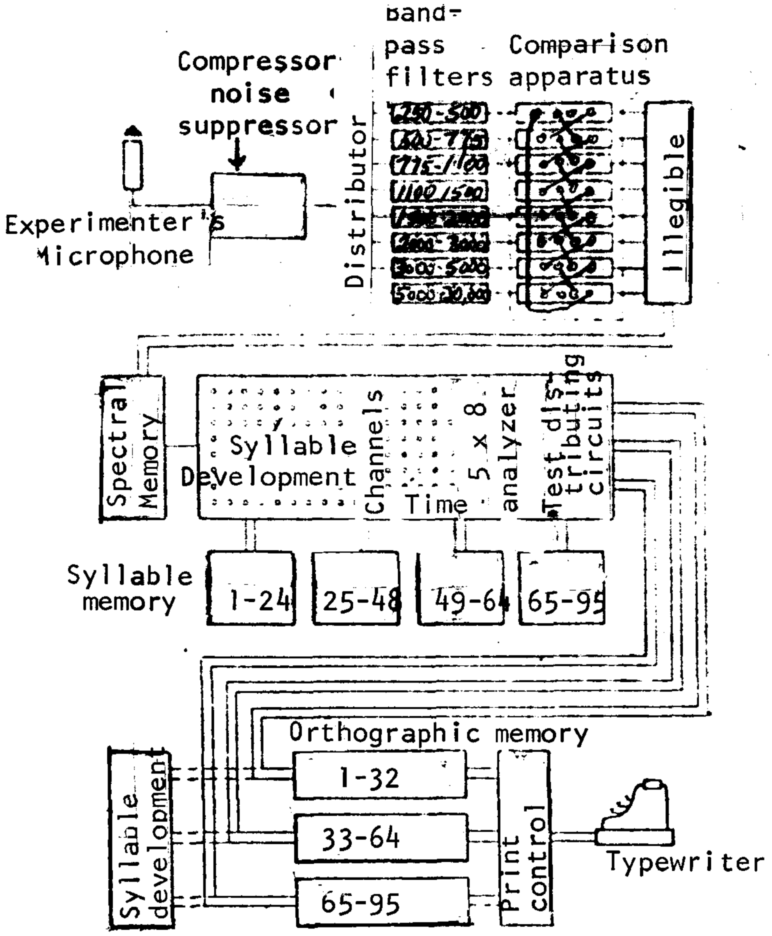

Figure 4. Block Diagram of Phonetic Typewriter of H. Olson and H. Belar

H. Olson and his colleagues developed a device in which recognition, recoding, printing and synthesis of speech is performed [125]. The memory of the machine consists of four English, eight French, four German and four Spanish words. The rate of operation of the machine is 60 words per minute. The accuracy of the operation with known speaker is 96-98%.

In the fourth model of the phonetic typewriter, still under development by J. Dreyfus-Graf [58], according to an announcement by the author, the specific features of the speech of various speakers have been eliminated. As in preceding models by this author, there is no memory, so that its accuracy of operation depends to a considerable extent on the carefulness of pronunciation of the speaker. Analysis of the speech spectrum is performed in ten filters (400-3200 Hz); also, there are other filters covering the 150-300 and 4500-6000 Hz range. Each spectral channel contains a detector and a low frequency filter for the determination of subformants (0-30 Hz) and a quantizer. The rate of change of the speech signal envelope is also determined. As a result, the speech signal is separated into 48 signals, each of which is quantized by time each one-fifteenth second and by three levels. The total information produced amounts to 1440 bit/sec, and the alphabet of the machine has 30 letters. No data on the accuracy of operation of the machine have yet been published.

The writing machine of M. Kalfaian [100] has a device for reducing the speech signal to a single base tone frequency using a generator controlled by the base tone of the input signal. They are filters whose outputs are compared after rectification with data contained in the memory by using electronic relays; the machine prints phonemes in a phonetic alphabet, not the ordinary written alphabet. No correction for errors in pronunciation is performed. The details of the operation of the machine have not yet been published.

The mechanical speech signal recognition machine developed at London University has, in contrast to the preceding machines, a linguistic memory. The inventor of the machine [51] states that the results of analysis can be used to control a typewriter, but he does not expect particularly great success. The recognition machine can distinguish four vowels and nine consonants. The differentiation is performed according to the maximum voltage from two of eighteen filters covering the speech signal range from 160 to 8000 Hz. It has been established, for example, that the vowel i corresponds to the maximum product of the outputs of the filters at 250 and 3200 Hz, the consonant m -- to the maximum product at 200 and 320 Hz. Consonants with almost identical spectra are distinguished further either according to length or voltage level. The linguistic memory contains information on the probability of sequences of two phonemes. The device consists of several potentiometers, and the information is input to the matrix of potentiometers by the position of their slide contacts. The acoustical device determines the form of the spectrum, length and intensity of sound and presence of base tone. The linguistic device contains standard combinations of phonemes present in speech, and combinations which do not exist in speech are forbidden. Thus, recognition occurs in the acoustical device and correction, i.e. improvement of intelligibility, occurs in the linguistic device. Experiments with 200 words have produced an accuracy of recognition of 72% (45% for any voice).

The correlation method of recognition is also based on multiplication of filter outputs. A device developed in the Bell Laboratories is designed for recognition of ten numbers pronounced by any subscriber [50]. The speech spectrum received is multiplied by a typical phoneme spectrum, after which the average value is determined. The maximum value for multiplication corresponds to the phoneme which is most similar to that received. The device consists of a complex containing ten relays, only one relay giving an output pulse for any one digit. The accuracy of recognition is 97-99% with a known speaker, 50-60% with any speaker. /388

In Japan, a device has been developed to recognize ten numbers pronounced separately. This process is performed using eight characteristics (number of voiced intervals per word, position of first and second formants at the beginning of the first voiced interval, position 100 msec after beginning of first voiced interval, etc.), with from two to thirteen levels each. The voltages corresponding to the levels of these eight characteristics are sent to a matrix circuit in which the conditional probabilities of all numbers are calculated. The highest conditional probability corresponds to the number actually pronounced. In pronunciation of one thousand words by one speaker, an accuracy of 99.7% was achieved, while when twenty different speakers (all men) pronounced one thousand words, an accuracy of 97.9% was achieved [189].

The results of an experimental investigation on the determination of various characteristics of numbers pronounced in Russian are presented in [30].

The time method of recognition is based on the usage of clipped speech. This method was presented in the works of I. Licklider [112] and subsequently in the works of other scientists [145, 150, 174], including Soviet scientists [12]. G. Tsemel' [28] uses clipped speech for differentiation of certain consonants. According to his data, the accuracy of recognition of the sounds p and t was 95%, the sound k -- 75%.

A. Rais [145] analyzed consonants based on their good differentiability by the ear. The experiments were performed to determine the capabilities for input of data in oral form to an automatic translation machine. It was established that the dependence of the number of pulses of clipped speech on the time for various vowels pronounced by different speakers is almost linear. The results of the experiments with three speakers are presented in Table 2. Preliminary differentiation of the signal was not performed.

TABLE 2

Speaker	a	o	u	i
	Pulses/sec			
first	558	450	350	283
second	533	417	317	--
third	491	384	292	218

I. Toffler suggested a method of separating the base tone from speech signals by using nonlinear elements and the clipping method [172].

Clipped speech was also used at Kyoto University [150]. Analysis was performed both for vowels and for consonants. According to the methodology used in this work, the speech signal was amplified to a preselected level, then fed to the input of a stage of Kipp oscillators, the outputs of which are pulses corresponding to the moments of changes in the direction of the rectangular clipped speech signal. The Kipp oscillator stage, the pulse lengths from which have different values in different cases, classify the clipped speech signal into fourteen values from 0.59-1.11 to 22.55-$32.22 \cdot 10^{-4}$ sec. In each channel, the number of pulses is determined by a counter. Analysis of vowels is performed according to the distribution of time intervals W(t) and the univariate distribution of probability $W_1(t)$ according to the formulas

$$W(\tau_{m_i}) = \frac{1}{\Delta\tau_i} \frac{\nu_i}{\sum\limits_{i=1}^{14} \nu_i} \text{ and } W_1(\tau_{m_i}) = \frac{1}{\Delta\tau_i} \frac{\tau_{m_i}\nu_i}{\sum\limits_{i=1}^{14} \tau_{m_i}\nu_i} \qquad (i = 1 \ldots 14).$$

where τ_{mi} is the mean time interval in the i-th channel; ν_i is the number of intervals in the i-th channel; $\Delta\tau_i$ is the time difference between the upper and lower boundaries of the i-th channel.

A writing machine based on analysis of clipped speech was constructed at the same university [149], and has a memory volume of 200 syllables. The details of operation of the machine have not yet been published. /389

A device for automatic distinction between twenty spoken words (the numbers from zero to nine, plus, minus, space, forward, back, etc.), based on clipping of preliminarily differentiated speech signals, was developed at the Academy of Sciences of the Georgian SSR [13]. It has been stated that the device is capable of distinguishing up to twenty spoken commands when tuned for a particular voice, or on the order of five commands from many voices.

The portable electronic calculating machine SHOEBOX, as yet the only series produced model, reacts to spoken pronunciation of the ten numbers (from zero to nine) and six additional commands: plus, minus, sum, partial sum, error, clear. The word "error" stops the device and clears all operations performed. Recognition occurs according to the location of the first phoneme, stressed vowel, last phoneme and time envelope of positive and negative peaks of the curve of the word. The operation of the device is unstable with different speakers [115].

Of the three writing machines which we have described [58, 123, 149], the most promising device, according to the materials of the Stockholm Seminar of 1962 [139] is the machine of J. Dreyfus-Graf, which has no limiting memory unit [32, 141].

Let us present certain other data concerning special devices for the analysis, recognition and usage of speech signals.

The analysis and recognition of audiofrequency signals can be performed using a device which has been developed, the operation of which is based on a resonant mechanical system consisting of glass fibers of various lengths and diameters up to 0.05 mm. One end of each fiber is fastened to a special shaped base, the other end can oscillate freely under the excitation of the audio waves. Using light beams, the oscillations of these free ends are projected through a standard screen onto photo diodes. The element consists of 2,000 fibers with a total volume of 16 cm^3, analyzing the frequency range from 30 to 20,000 Hz with a resolving capacity of about 10 Hz. Each standard screen stores the signals of prototypes. The photo element integrates the entire quantity of light, and the more closely the signal being analyzed corresponds to the standard signal represented by the standard screen, the greater the total light flux. If the flux exceeds a given threshold, the signal is recognized. Since the standard can also be changed in correspondence to the information produced, this element has characteristics of self-teaching. The device recognizes only short words pronounced by a given voice. This device was constructed in an attempt to establish communications with dolphins. It was established that the "speech" of dolphins consists of short signals (approximately 0.1 sec) at frequencies from 5 to 10 KHz [180, 181].

Harmonic analysis of periodic functions fixed in the form of graphs or tables can be performed using an electromechanical harmonic analyzer allowing simultaneous production of five pairs of coefficients of Fourier series with an accuracy to 0.3% of the maximum value of the function being analyzed [4].

A device has been constructed in which the recognition of commands (numbers) is achieved using visual data on the movement of the lips. Light sources with directed reflectors are installed on each lip. A photo resistor with a collecting reflector is placed before the lips, connected to one arm of a balanced bridge. The voltage taken from the bridge is amplified, fed to a differential amplifier and thence to a strip chart recorder. The light sources and photo resistors with reflector are connected to the head of the speaker. The intelligibility of ten numbers for a concrete speaker was 91%, for two different speakers -- 78.3%. By determining one more parameter -- the rate of movement of the air flow near the lips -- the intelligibility for a single speaker could be increased to 100%, for two speakers -- to 81% [84].

A system of solenoids has been theoretically developed, using which different codes can be produced for 24,000 English words up to 16 letters long; however, this system for word recognition has not yet been constructed. Essentially, the system is a memory device in which the curves of speech signals are stored in digital form and which can be used to produce momentary values of the correlation function of the speech signal [38, 134].

Based on the usage of relay systems, a device has been created for conversion of digital information to speech information. The sounds of the numbers zero to nine and the words "volt," "second," "power" are recorded on the tracks of a magnetic drum. The address of each drum track is output by a ring counter, and the recording thus selected is sent to an audio amplifier [144]. /390

The neuron network and auditory apparatus of man have been modeled by several authors [19, 74, 83, 90]; also, a correlation theory of hearing has been developed [101]. On the basis of this model, a device has been developed for the recognition of sounds and individual words. The device consists of an output amplifier, a system of filters and a unit of logic and computer circuits modeling the outer and middle ear, the inner ear and the nerve networks respectively. The voiced consonants b, d, g and unvoiced consonants p, t and k are differentiated using differentiating logic circuits; however, no complete electronic analog has yet been produced [116].

Experiments have been performed in the "understanding" of speech which is not heard. Information on a phoneme is transmitted through the hand using 24 vibrators connected to the outputs of a functional model of the auditory organ. The numbers from one to nine were recorded on magnetic tape. The experimental subject is told if he makes an incorrect answer, and the number is repeated. After two to three hours of training, 85% correct answers were produced. This experiment is of great significance for the deaf [190].

The human voice is not symmetrical about a transitional axis as is, for example, the noise spectrum. This peculiarity, "asymmetry of the envelope," was used for the creation of a safety switch which turns off a powerful machine, such as a machine tool, if the operator shouts [163].

A speech signal analyzer has been developed, consisting of 54 gaussian type filters [80]. Up to 1000 Hz, the filters have a width of 70 Hz each, while over 1000 Hz the width increases by 6.5% each step. The output of each filter is rectified and quantized; the quantized currents can be tracked visually and, after the proper processing, can be input to an electronic computer [169].

A speech signal analyzer has been developed which consists of 96 filters and covers the frequency range from 30 to 8,000 Hz. The signals are normalized, detected, up to the fourth derivative is determined, and the output of the analyzer is connected to a triple beam oscilloscope. The process of recognition is currently being automated [53].

The solution of the problem of separating the base tone is of great significance both for linguists and for the manufacturer of vocoders and for solution of the problem of recognition of speech signals in general [78]. Several variants of a special apparatus [42, 56, 147] and devices have been developed which have units for determining the autocorrelation function of signals in order to separate peak maxima [72], the widths of formants [62, 170] and other supplementary devices [143, 184]. For example, in [184] it is suggested that the phase of the outputs of N filters be changed by 90 degrees. These outputs are looked upon as multidimensional vectors which change with time. If there is periodicity in the signal, the vector will form a closed curve. Five filters of 120 Hz width each are used, covering the frequency range from 300 to 900 Hz.

In order to separate the base tone in a speech signal, a device has been developed consisting of elements with nonlinear characteristics and slight time constants [151]. Another system for analysis of the frequencies of formants and base tone operates on the principle of the tracking filter with preliminary transfer of the spectrum of frequencies being analyzed [32] or generation of a signal with manual adjustment of the signal to correspond to the signal being analyzed [179].

5. Universal Computers As Means of Investigating and Recognizing Speech Signals /391

The usage of universal computers for the investigation of speech signals began shortly after computers were invented [48, 70, 83] and has expanded each year. They are used in connection with spectral analysis of speech signals [5, 37, 46, 99], for recognition and synthesis of a selected phoneme combination [38, 88, 91, 116], or numbers [52, 71, 152], to separate the base tone [46, 72, 81, 120] and determine the parameters of formants [127, 167, 182], to investigate the operation of vocoders [132], to determine the possibilities for input of speech signals into computers for mathematical modeling of communications systems [15, 46], etc. In [120], a method is presented for expansion of a speech signal into a Fourier series and determination of its constants. The amplitudes of each sequence of frequencies are logarithmized and analyzed in a second spectral analyzer. The output of this analyzer is the logarithm of the power spectrum and has peak values in the case of analysis of voiced phonemes, and has no peak values if the phonemes are unvoiced or have no base tone. Since the time of changes of the frequencies to the speech signals cause periodic pulsations in the amplitude spectrum, the Fourier transform of the spectrum gives the frequency of the pulsations, inversely proportional to the frequency of the base tone of the speech.

Spectral analysis has been performed for a group of preliminarily segmented vowel sounds, using a special device which performed digital coding of the speech segments being analyzed. These data were input to the computer and a stepped synchronous analysis was performed using the Fourier method [37].

In [135], a new technique is suggested for measurement of the frequency and widths of formants of a speech signal, based on the theory of Fant [24]. A form is assigned to the spectral equations:

$$f(t) = a_0 + \sum_{i=1}^{N} e^{-\pi B_i t} (a_i \cos 2\pi F_i t + b_i \sin 2\pi F_i t),$$

where N = 3 and 4 (number of formants).

A method is presented for determination of the numerical values of the coefficients a_0, a_i, b_i, F_i ($1 \leq i \leq N$) by computer using the least squares method (B_i is the width of the i-th formant and F_i is its frequency). Based on the experimental materials from analysis of three words (bought, bottle and beet) pronounced by two persons twice each, it is affirmed that the first two formants for ə, i and a are found quite reliably.

The energy spectrum of speech signals can be calculated by methods other than expansion into Fourier series, for example by separation of this spectrum into polynomials or representation of the spectrum as a Markov process [98]. In [5], the results are presented from a correlation-spectral analysis, while [10] presents estimates of the error in the spectral analysis method and [7] presents an estimate of the frequency of quantization of the speech spectrum with correlation and spectral analysis by computer.

It is stated in [148] that the problem of recognition of various patterns, including speech patterns, can be reduced to the problem of finding a class in which a given signal belongs if the total number of classes of signals in which the signal may be included is known. This problem is solved by minimizing a certain risk function, as a result of which the optimal rules are found to be used for solving the problem of recognizing output signals from an electronic analyzer, a volume of the auditory helix.

A method has been suggested for representing a model of a formant as an n-dimensional vector, each component of which is a single discrete value of the formant at a given moment in time [40]. Thus, the dimensions of an n-dimensional space are determined by multiplying the discrete values of parameters of the formant by the moments of observation. The computer develops this data in two steps, called by the inventor of the process learning and recognition. The results of analysis of each of ten phonemes /392 (first letters of the alphabet) pronounced by ten speakers, are presented in the form of a matrix [40].

It has been suggested that a computer program be developed to calculate the energy and envelope of the speech signal over a fixed time interval, the frequency of transition of the signal through the zero level and the distribution of intervals between zeros in clipped speech throughout the entire time interval, the autocorrelation and mutual correlation functions and also to perform spectral analysis of the speech signal in order to recognize the signals [9]. An analog-digital converter with eight digit readout of vinary numbers has been developed for this purpose for input of sound information into the computer [6].

In [53], a speech signal is analyzed in a 30 channel band type analyzer. The computer is used to determine the frequencies F_1, F_2 and F_3 each 10 msec. The sectors of speech during which the formant frequencies do not change are not taken into consideration. An experiment for the recognition of ten monosyllabic words spoken three times each by three speakers gave 100% correct results. Recognition of speakers by their voices was also successful.

The problem of separating the base tone has also been the subject of many works. In addition to special apparatus outlined above, this work is being performed by computer, in most cases together with determination of other speech parameters. In [81, 182], a program is developed which is used to take a speech signal preliminarily processed in a correlation type analyzer through an analog-digital converter, introduce it into a computer for determination of the parameters of the base tone and other indicators of the speech signal such as: frequency and amplitude of the first three formants, instantaneous signal power and rate of change of all quantities. The results of separation of the base tone are compared to data produced in [109], in which the determination of the base tone and rate of its change during the pronunciation of individual words· and syllables was performed generally automatically, while more precise determination was made by manual measurement of distances between peaks on oscillographs of the speech signals. Good correspondence of the measurement results is noted.

In [167-169], a process is analyzed for separating the frequencies of formants and determining the formants in one-syllable words of the Japanese language by computer. First, the speech is preliminarily analyzed in a filter system covering the frequency range from 200 to 5900 Hz, then the speech is coded by an eight bit binary word and sent to a computer magnetic tape recorder. In the computer, the formant frequencies are separated from the speech, the rate of change of formant frequencies is determined and the second order moment is defined near the mean frequency; the phonemes are segmented and phoneme classification of the spectrum is performed. It is stated that this method allows error free determination of separately pronounced phonemes, as well as almost error free determination and recognition of sounds and unvoiced consonants.

In order to automate the exchange of information between a man and a warehouse, a system has been created in which audible speech is analyzed according to certain energy characteristics, converted into binary electrical signals using a device containing frequency filters, analog-digital converters and decoders, is coded onto punched cards and transmitted to an electronic computer. A subscriber can access to the computer by telephone and receive an answer in various languages (French, English, etc.) [107]. A similar system (using the IBM-7770 computer) with a memory volume of 60 words can answer 750 simultaneous telephone inquiries concerning prices on an exchange [8].

A unique trend in the investigation of speech signals by computer, resulting from the search for methods to reduce the quantity of information concerning the sound of speech, has appeared in the usage of the "analysis-synthesis" method, suggested by K. Stevens et al. [86, 98, 127]. In [127], the initial determination of speech signal parameters is performed by comparison of the input spectrum with a spectrum formed in the system as a result of combination of six curves for the first formants and six for the second formants, i.e. 36 standard spectral curves in all. Eight parameters are changed in forming the standard spectra: the frequency and width of the bands of the first three formants, the frequency of the fourth formant and the position of the zero in the source spectrum. It has been stated that this method allows sufficiently precise investigation of speech and the production of initial data for the development of speech recognition apparatus.

In another suggestion, the speech signal is subjected directly to mathematical analysis in the computer without supplementary apparatus and is approximated by 30 orthogonal functions in the form of an exponentially damped Fourier series. When the sound is changed, the numerical values of the coefficients are changed, while the functions themselves remain unchanged. The symbols produced are used for synthesis [55]. /393

Similar works are also being conducted in Japan [91, 98] and the Polish People's Republic [97].

A decrease in the influence of the subjective properties of the speakers on the result of machine recognition can be achieved to a certain extent by using the principle of self-tuning and self-teaching. If many people are talking, as in an ordinary discussion, this principle is, of course, ineffective. In [173], the results are described of a work on the recognition of patterns using these methods in combination with "analysis-synthesis" methods. Using 20 characteristics, for example, a great quantity of combinations can be produced, but many of these characteristics are nonessential, so that submatrices are formed including only the essential characteristics. The degree of importance of each characteristic and combination of characteristics is determined during the course of self-teaching, i.e. by synthesis of the required "dictionary." At first, the machine generates all characteristics and, analyzing the relationships between individual cells, determines the weight of each one, generating new relationships and determining their weights in turn. Recognition is performed by the comparison method. The number of proper answers achieved by recognizing known stylized portraits and spectrograms of speech was 80-100%, for unknown portraits and spectrograms 60-100%.

An algorithm for machine recognition using elements of self-tuning was also developed in [183]. With a memory volume of eighteen words, the accuracy of differentiation achieved by this self-tuning system in processing information input to the machine in four languages was 96%; when twenty words were used, the accuracy was 86%. The system could successfully distinguish a voice concerning which no preceding experience had been accumulated.

A program was developed which, with a limited memory volume (numbers from zero to nine, plus, minus, equals, parentheses, etc., 83 words in all) the computer "learned" to recognize these words both for a known and for a random speaker (although with poorer results), to perform arithmetic operations on command with the number introduced by voice, to print out the results and to translate them into another language [132].

A computer machine for the recognition of speech signals not only has the advantage that it makes it possible to use partial information and algorithms developed for the machine for translation from one language to another, but also has the advantage that its large memory volume and high operating speed allow speech signals to be separated for analysis into extremely short amplitude and frequency sectors, and allows data to be compared with standard data quite rapidly. On the other hand, in order to achieve universality of recognition, i.e. independence of the results of analysis from subjective properties of the speakers, it should not contain comparison elements; also, large computers are expensive and cannot be mobile. Therefore, in spite of certain advantages over special machines, the latter, i.e. direct analysis machines, are more promising for the final solution of the problem of machine recognition of speech signals and usage of these signals in various control and communications systems.

In this article we have touched upon the most general part of the work in the area of machine recognition and have not discussed the problem of speech synthesis at all. The main factor reducing the accuracy of the operation of speech recognition apparatus is the inconstancy of the formant indicators, depending to a great extent on the individualities of the speakers, and the difficulty of segmenting. However, the results of investigations have already been used in communications technology, as well as military affairs. For example, vocoders have been developed in the USA which are used in aviation [2, 133]. As the apparatus is improved, the area of their application will doubtless increase, as a result of which cybernetic systems will receive new elements which will react to spoken commands without intermediate coding.

REFERENCES

1. "Analysis of Man's Behavior by His Speech," *Elektronika*, vol. 37, No. 9, 1964. /394
2. "The Army Orders a Miniature Vocoder," *Elektronika*, vol. 37, No. 15, 1964.
3. Varshavskiy, L. A., I. M. Litvak, "Investigation of Formant Composition and Certain Other Physical Characteristics of the Sounds of Russian Speech," *Problemy Fiziol. Akustiki*, No. 3, 1955.
4. Vasilenko, A. T., Yu. N. Denisov, "An Electromechanical Harmonics Analyzer," *Pribory i Tekhnika Eksperimenta*, No. 6, 1963.
5. Voloshin, G. Ya., "Spectral Analysis of Speech Signals by Electronic Computer," *Sbornik Trudov Inta Matematiki SO AN SSSR, Vychislitel'nyye Sistemy* [Collected Works of Institute of Mathematics, Siberian Department Academy of Sciences USSR, Computer Systems], No. 10, Novosibirsk, 1964.
6. Voloshin, G. Ya., "Analog-Digital Converter for Input of Speech Signals to Electronic Computers," *Sbornik Trudov Inta Matematiki SO AN SSSR, Vychislitel'nyye Sistemy* [Collected Works of Institute of Mathematics, Siberian Department Academy of Sciences USSR, Computer Systems], No. 10, Novosibirsk, 1964.
7. Voloshin, G. Ya., "The Frequency of Sampling of a Random Function with Correlation-Spectral Analysis," *Sbornik Trudov Inta Matematiki SO AN SSSR, Vychislitel'nyye Sistemy* [Collected Works of Institute of Mathematics, Siberian Department Academy of Sciences USSR, Computer Systems], No. 14, Novosibirsk, 1964.
8. "The Computer Answers the Telephone," *Elektronika*, vol. 37, No. 5, 1964.

9. Zagoruyko, N. G., "The Exchange of Oral Information Between Man and Computer Systems," *Sbornik Trudov Inta Matematiki SO AN SSSR, Vychislitel'nyye Sistemy* [Collected Works of Institute of Mathematics, Siberian Department Academy of Sciences USSR, Computer Systems], No. 10, Novosibirsk, 1964.

10. Zagoruyko, N. G., "Error in Calculating Energy and Envelope of Speech Signal by Computer," *Sbornik Trudov Inta Matematiki SO AN SSSR, Vychislitel'nyye Sistemy* [Collected Works of Institute of Mathematics, Siberian Department Academy of Sciences USSR, Computer Systems], No. 10, Novosibirsk, 1964.

11. Zagoruyko, N. G., G. Ya. Voloshin, V. N. Yelkina, "Automatic Recognition of Speech Patterns," *Sbornik Trudov Inta Matematiki SO AN SSSR, Vychislitel'nyye Sistemy* [Collected Works of Institute of Mathematics, Siberian Department Academy of Sciences USSR, Computer Systems], No. 14, Novosibirsk, 1964.

12. Kakauridze, A. G., "Some Problems of Coding of Speech Vowel Sounds," *Trudy Inta Elektroniki, Avtomatiki i Telemekhaniki AN GruzSSR* [Works of the Institute of Electronics, Automation and Telemechanics, Academy of Sciences Georgian SSR], No. 1, 1960.

13. Kakauridze, A. G., "Experimental Device for Automatically Distinguishing Among a Limited Set of Speech Commands," *Elementy Vychislitel'noy Tekhniki i Mashinnyy Perevod, In-t. Elektroniki, Avtomatiki i Telemekhaniki AN GruzSSR* [Elements of Computer Equipment and Machine Translation, Institute of Electronics, Automation and Telemechanics, Academy of Sciences Georgian SSR], Tbilisi, 1964.

14. Caldwell, V., E. Glasser, D. Stewart, "An Analog Model of the Ear," *Sbornik Problemy Bioniki* [Collection of Problems of Bionics], Moscow, 1965.

15. Lebman, Yu. A., V. N. Sobolev, "Analog-Digital Converter for Input of Speech Signal to Computer Machine," *Elektrosvyaz'*, No. 8, 1963.

16. Myasnikov, L. L., "Objective Recognition of Speech Sounds," *Zh. Tekhnich. Fiz.*, vol. 13, No. 3, 1943.

17. Myasnikov, L. L., "The Sounds of Speech and Their Objective Recognition," *Vestnik Leningr. Universiteta*, No. 3, 1946.

18. Myasnikov, L. L., "Physical Investigation of the Sounds of Russian Speech," *Izvestiya Akad. Nauk SSSR, Seriya Fiz.*, v. 13, No. 6, 1949.

19. Myuler, P., T. Martin, F. Puttsrat, "General Principles of Operation in Neuron Networks and Their Application to the Recognition of Acoustical Patterns," *Sbornik Problemy Bioniki*, Moscow, 1965.

20. Pirogov, A. A., "Theoretical Considerations Concerning a Method for Coding and Synthesis of Speech Information Using a Harmonic Function," *Dokl. na Vsesoyuzn. Soveshch. Sektsii Rechi Komissii po Akustike AN SSSR* [Report at All-Union Conference of Speech Section of Commission on Acoustics, Academy of Sciences USSR], Moscow, 1958.

21. Pirogov, A. A., "Harmonic System for Compression of Speech Spectra," *Elektrosvyaz'*, No. 3, 1959.

22. Sapozhkov, M. A., *Rechevoy Signal v Kibernetike i Svyazi* [The Speech Signal in Cybernetics and Communications], Moscow, 1963.

23. Solodovnikov, V. V., *Statisticheskaya Dinamika Linenykh Sistem Avtomaticheskogo Upravleniya* [Statistical Dynamics of Linear Automatic Control

Systems], Moscow, 1960.

24. Fant, G., *Akusticheskaya Teoriya Recheobrazovaniya* [The Acoustical Theory of Speech Formation], Moscow, 1964.
25. Kharkevich, A. A., *Spektry i Analiz* [Spectra and Analysis], Moscow, 1953.
26. Kharkevich, A. A., *Ocherki Obshchey Teorii Svyazi* [Essays on the General Theory of Communications], Moscow, 1955.
27. Khrapovistkiy, A. V., "Theoretical and Experimental Investigation of a Cosine-Logarithmic Vocoder," *Dokl. na Vsesoyuzn. Soveshch. Sektsii Rechi Komissii po Akustike AN SSSR* [Report at All-Union Conference of Speech Section of Commission on Acoustics, Academy of Sciences USSR], Moscow, 1959.
28. Tsemel', G. I., "Determination of the Invariant Characteristics of Plosive Sounds on the Basis of Clipped Speech Signals," *Izvestiya Akad. Nauk SSSR Otdel Tekhnicheskikh Nauk Energetika i Avtomatika*, No. 4, 1959.
29. Tsemel', G. I., "Automatic Recognition of Speech Sounds," *Zarubezhnaya Radioelektronika*, No. 4, 1962.
30. Tsemel', G. I., "Recognition of a Small Group of Words by Characteristic Features of the Speech Signal," *Sbornik Problemy Peredachi Informatsii*, No. 16, Moscow, 1964.
31. Chistovich, L. A., "Influence of Frequency Limitations on Intelligibility /395 of Russian Consonant Sounds," *Dokl. na Vsesoyuzn. Soveshch. Sektsii Rechi Komissii po Akustike AN SSSR* [Report at All-Union Conference of Speech Section of Commission on Acoustics, Academy of Sciences USSR], Moscow, 1956.
32. Abstracts of Papers on Speech Analysis, Stockholm Speech Communications Seminar, 1962, Royal Institute on Technology, Stockholm, Sweden, *The Journal of the Acoustical Society of America (=JASA)*, Vol. 35, No. 7, 1963.
33. Bennett, W. R., "The Correlatograph. A Machine for Continuous Display of Short Term Correlation." *Bell System Techn. J.*, Vol. 32, Sept. 1953.
34. Billings, A.R., "Simple Multiplex Vocoder," *Electronic and Radio Engr*, Vol. 36, No. 5, 1959.
35. Billings, A.R., "Communication Efficiency of Vocoders Comparison of Low-Power and Conventional Systems," *Electronic and Radio Engr*, Vol. 36, No. 12, 1959.
36. Bogert, B. P., "Vobanc - a Two-to-One Speech Band-Width Reduction System," *JASA*, Vol. 28, No. 3, 1956.
37. Borenstein, D.P., "Spectral Characteristics of Digit-Simulating Speech Sounds," *Bell System Techn, J.*, Vol. 42, No. 6, 1963.
38. Brick, D.B., and G. G. Pick, "Microsecond Word-Recognition System," *IEEE Trans. Electronic Comput.* EC-13, No. 1, 1964.
39. Campanella, S.A., "A Survey of Speech Bandwidth Compression Techniques," *IRE Trans. Audio*, AU-6, No. 5, 1958.
40. Campanella, S. I., Coulter, D. C., and P. Engler, "Speech Recognition by Formant Pattern Matching in N-dimensional Space," *JASA*, Vol. 36, No. 5, 1964.
41. Campanella, S. I., Coulter, D. C., and R. Irons, "Influence of Transmission Error on Formant Coded Compressed Speech Signals," *J. Audio Engng. Soc.*, Vol. 19, No. 2, 1962.

42. Carre, R., Lancia, R., Paille, J., and R. Gsell, "Etude et realisation d'un detecteur de melodie pour analyse de la parole," *Onde electr.*, Vol. 43, No. 434, 1963.
43. Chang, S. H., "Two Schemes of Speech Compression System," *JASA*, Vol. 28, No. 4, 1956.
44. Chang, S. H., Pihl, C. E., and I. Wiren, "The Intervalgram as a Visual Representation of Speech Sounds," *JASA*, Vol. 23, No. 6, 1951.
45. Cherry, C., Halle, M., and R. Jacobson, "Toward a Logical Description of Languages in Their Phonemic Aspects," *Language*, Vol. 29, No. 1, 1953.
46. Clapper, C. L., "Digital Circuit Techniques for Speech Analysis," *IEEE Trans. Communication and Electronics*, CE-11, No. 66, 1963.
47. Cramer, B., "Sprachsynthese zur Obertragung mit sehr geringer Kanalkapazitat, *Nachrichtentechn.* Z., No. 8, 1964.
48. David, E. E., and H. S. McDonald, "Note on Pitch-Synchronous Processing of Speech," *JASA*, Vol. 28, No. 6, 1956.
49. David, J., Schroeder, M. K., Logan, B. F., and H. J. Prestigiacomo, "Voice-Excited Vocoders for Practical Speech Bandwidth Reduction," *IRE Trans. Inform. Theory*, IT-8, No. 5, 1962.
50. Davis, K. H., Bidulph, R., and S. Balashek, "Automatic Recognition of Spoken Digits," *JASA*, Vol. 24, No. 6, 1952.
51. Denes, P., "The Design and Operation of the Mechanical Speech Recognizer at University College London," *J. Brit. IRE*, Vol. 19, No. 4, 1959.
52. Denes, P., and M. V. Mathews, "Spoken Digit Recognition Using Time-Frequency Pattern Matching," *JASA*, Vol. 32, No. 11, 1960.
53. Deuber, G., "Varied Approaches Used to Develop System for Reliable Recognition of Voice Commands," *Electronic News*, Vol. 8, No. 383, 1963.
54. Deweze, A., "Techniques des reconnaissance automatique des formes visuelles et sonores," *Automatisme*, Vol. 9, No. 3, 1964.
55. Dersch, W. C., "Speech Operated Safety Switch," *Electronics*, Vol. 36, No. 25, 1963.
56. Dolansky, I. O., "Instantaneous Pitch-Period Indicator," *JASA*, Vol. 27, No. 1, 1955.
57. Dreyfus-Graf, J., "Phonetographe et Subformants," *Bull. Techn. PTT, Bern*, No. 2, 1957.
58. Dreyfus-Graf, J., "Phonetographe: Present et Future," *Bull. Techn. PTT, Bern*, No. 5, 1961.
59. Dudley, H. W., "Remaking Speech," *JASA*, Vol. 11, No. 2, 1939.
60. Dudley, H. W., "Speech Analysis and Synthesis System," *JASA*, Vol. 22, No. 6, 1950.
61. Dudley, H. W., "Phonetic Pattern Recognition for Narrowband Transmission," *JASA*, Vol. 30, No. 8, 1958.
62. Dunn, H. K., "Methods of Measuring Vowel Formant Bandwidths," *JASA*, Vol. 33, No. 12, 1961.
63. Fano, R. M., "The Information Theory Point of View in Speech Communication," *JASA*, Vol. 22, No. 10, 1950.
64. Fant, G., "Acoustic Theory of Speech Production with Calculations Based on X-Ray Studies of Russian Articulations," *Mouton & Co, s'Gravenhage*, 1960.
65. Fant, G., Fintoft, K., Liliencrants, J., Lindblom, B., and J. Martony, "Formant-Amplitude Measurements," *JASA*, Vol. 35, No. 11, 1963. /396
66. Flanagan, I. L., "A Difference Limen for Vowel Formant Frequency," *JASA*, Vol. 27, No. 3, 1955.

67. Flanagan, J. L., "Automatic Extraction of Formant Frequences from Continuous Speech," *JASA*, Vol. 28, No. 1, 1956.
68. Flanagan, J. L., "Band-Width and Channel Capacity Necessary to Transmit the Formant Information of Speech," *JASA*, Vol. 28, No. 4, 1956.
69. Flanagan, J. L., "A Resonance-Vocoder and Baseband Complement: A Hybrid System for Speech Transmission," *IRE. Trans. Audio*, AU-8, May-June 1960.
70. Forgie, J. W., and C. W. Hughes, "A Real-Time Speech Input System for a Digital Computer," *JASA*, Vol. 30, No. 7, 1958.
71. Forgie, J. W. and C. D. Forgie, "Results Obtained from a Vowel Recognition Computer Programme," *JASA*, Vol. 31, No. 9, 1959.
72. Fujisaki, H., "Automatic Extraction of Fundamental Period of Speech by Auto-Correlation Analysis and Peak Detaction," *JASA*, Vol. 32, No. 11, 1960.
73. Gabor, D., "New Possibilities in Speech Transmission," *J. IRE*, Vol. 94, Nov. 1948.
74. Galdwell, W. F., "Recognition of Sounds by Cochlear Patterns," *IEEE Trans. Military Electronics*, MIL-7, No. 2-3, 1963.
75. Gerstman, L. J., Liberman, A. M. Delattre, P. C., and F. S. Cooper, "Rate and Duration of Change in Formant Frequency as Cues for Identification of Speech Sounds," *JASA*, Vol. 26, No. 9, 1954.
76. Gold, B., and C. Rader, "Bandpass Compressor: A New Type of Speech-Compression Device," *JASA*, Vol. 36, No. 6, 1964.
77. Golden, R. M., MacLean D. I., and A. I. Prestigiacomo, "Frequency Multiplex System for a 10-Spectrum-Channel Voice-Excited Vocoder," *JASA*, Vol. 36, No. 10, 1964.
78. Gribenski, A., "The Pitch of Sound, Its Measuring and Perception," *Nature*, Vol. 173, No. 4, 1957.
79. Haggard, M. P., "In Defense of the Formant," *Phonetica*, Vol. 10, No. 3-4, 1963.
80. Harris, C. M., and W. M. Waite, "Gaussian-Filter Spectrum Analyzer," Vol. 35, No. 4, 1963.
81, Harris, C. M. and M. R. Weiss, "Pitch Extraction by Computer Processing of High-Resolution Fourier Analysis Data," *JASA*, Vol. 35, No. 3, 1963.
82. Hellwarth, G. A., "Speech-Formant Measurement with a Continuously Tuned Automatic Tracking Filters," *JASA*, Vol. 35, No. 5, 1963.
83. Heydeman, P., "Ein Modellversuch zum Frequenzunterscheidungsvermogen des Ohres," *Acustica*, Vol. 13, No. 2, 1963.
84. Hillix, W., "Use of Two Nonacoustic Measures in Computer Recognition of Spoken Digits," *JASA*, Vol. 35, No. 12, 1963.
85. Howard, C. R., "Speech Analysis Synthesis Scheme Using Continuous Parameter," *JASA*, Vol. 28, No. 6, 1956.
86. Howard, C. R., Chang, S. H., and M. J. Carrabes, "Analysis and Synthesis of Formants and Moments of Speech Spectra," *JASA*, Vol. 28, No. 4, 1956.
87. Huggins, W. H., "A Phase Principle for Complex-Frequency Analysis and its Implications in Auditory Theory," *JASA*, Vol. 24, No. 6, 1952.
88. Hughes, G. W., "Identification of Speech Sounds by Means of a Digital Computer," *JASA*, Vol. 31, No. 1, 1959.
89. Husson, R., "Zur Spektraistruktur menschlicher Vokale aller Stimmstarken," *Phonetica*, No. 1-2, 1963.
90. Inomata, S., "An Auditory Pattern Processing Model," *IEEE Trans. Inform. Theory*, IT-9, No. 4, 1963.

91. Inomata, S., "Speech Recognition and Generation by a Digital Computer," *Res. Electrotechn. Labs,* No. 645, 1963.

92. Ithell, A. H., "A Determination of the Acoustical Input Impedance Characteristics of Human Ears," *Acustica,* Vol. 13, No. 4, 1963.

93. Jaffe, J., Cassotta, L., and S. Feldstein, "Markovian Model of Time Patterns of Speech," *Science,* Vol. 144, No. 3260, 1964.

94. Jakobson, R., Fant, G. G., and M. Halle, "Preliminaries to Speech Analysis, The Distinctive Features and Their Correlates," *Massachusetts Institute of Technology. Acoust. Lab. Techn. Rept. No. 13,* 1952.

95. Jakobson, R., "Die Verteilung der stimmhaften und stimmlosen Gerauschlaufe im Russischen, Festschrift fur Max Vasmer," *Berlin,* 1956.

96. Johnson, W., "System to Generate Speech from Written Pattern Shown," *Electronic News,* Vol. 8, No. 392, 1963.

97. Kacprowski, J., "Speech Compression by Means of Analysis-Synthesis Methods," Polish Academy of Sciences, *Proc. Vibration Probl.,* Vol. 5, No. 3, 1964.

98. Kadokava, J. and K. Nakata, "Formant Frequency Extraction by Analysis-by- /397
Synthesis Technique," *J. Radio Res. Labs,* Vol. 10, No. 49, 1963.

99. Kadokawa, J., and K. Nakata, "Analysis of Speech by Vocal Tract Configuration," *J. Radio Res. Labs,* Vol. 11, No. 54, 1964.

100. Kaliaian, M. V., "Phonetic Typewriter of Speech," *JASA,* Vol. 36, No. 6, 1964.

101. Karplus, H. B., "Correlation Hypothesis to Explain the Fine Frequency Discrimination of the Ear," *JASA,* Vol. 35, No. 5, 1963.

102. Klass, P. I., "Vocoder Increases Channels Security," *Aviat. Week,* Vol. 73, No. 4, 1960.

103. Klumpp, R. G., and I. C. Webster, "Intelligibility of Time-Compressed Speech," *JASA,* Vol. 33, No. 3, 1961.

104. Loenig, W., "A New Frequency Scale for Acoustic Measurements," *Bell Labs, Rec.,* Vol. 27, No. 7, 1949.

105. Ladefoged, P., "Acoustic Correlate of Subglottal Activity," *JASA,* Vol. 35, No. 5, 1963.

106. Ladefoged, P., and N. P. McKinney, "Loudness, Sound Pressure and Subglottal Pressure in Speech," *JASA,* Vol. 35, No. 4, 1963.

107. Latil de, P., "La parole est aux calculateurs," *Electronique Industr.* No. 76, 1964.

108. Lehiste, I., and G. E. Peterson, "Some Basic Considerations in the Analysis of Intonation," *JASA,* Vol. 33, No. 4, 1961.

109. Lieberman, P., "Perturbations in Vocal Pitch," *JASA,* Vol. 33, No. 5, 1961.

110. Liberman, A. M., Ingeman, F., Lisker, L., Delattre, P. C., and F. S. Cooper, "Minimal Rules for Synthesizing Speech," *JASA,* Vol. 31, No. 11, 1959.

111. Licklider, I. C., "Effects of Amplitude Distortion Upon the Intelligibility of Speech," *JASA,* Vol. 18, No. 2, 1964.

112. Licklider, I. C., "The Intelligibility of Amplitude-Dichotomized. Time-Quantized Speech Waves," *JASA,* Vol. 22, No. 6, 1950.

113. Licklider, I. C., "Influence of Phase Coherence Upon the Pitch of Complex Periodic Sounds," *JASA,* Vol. 27, No. 5, 1955.

114. Licklider, I. C., "Man-Computer Symbiosis," *IRE Trans.* HFL-1, No. 1, 1960.

115. "Machines Controlled by Spoken Commands," *Datamation,* Vol. 8, No. 6, 1962.

116. Martin, T. B., and I. I. Talvagel, "Application of Neural Logic to Speech Analysis and Recognition," *IEEE Trans. Military Electronics,* MIL-7, No. 2-3, 1963.

117. Miller, A. E., and M. V. Mathews, "Investigation of the Glottal Waveshape by Automatic Inverse Filtering," *JASA*, Vol. 35, No. 11, 1963.

118. Miller, R. L., "Improvements in the Vocoder," *JASA*, Vol. 25, No. 4, 1953.

119. Nicolau, E., Weber, L., and S. Gavat, "Aparate pentru recunoasterca automata a vocalelor," *Automatica ci electronica*, Vol. 7, No. 6, 1963.

120. Noll, A. M., "Short-Time Spectrum and 'Cepstrum' Techniques for Vocal-Pitch Detection," *JASA*, Vol. 36, No. 2, 1964.

121. Oeren, F. W., "Some Critical Observation on the Formant Theory of Vowel Recognition," *Phonetica*, Vol. 10, No. 1-2, 1963.

122. Olson, H. F., and H. Belar, "Phonetic Typewriter," *JASA*, Vol. 28, No. 6, 1956.

123. Olson, H. F., and H. Belar, "Phonetic Typewriter III," *JASA*, Vol. 33, No. 11, 1961.

124. Olson, H. F., and H. Belar, "Syllable Analyzer, Coder and Synthesizer for the Transmission of Speech," *IRE Trans. Audio*, AU-10, No. 11, 1962.

125. Olson, H. F., Belar, H., and R. Sobrino, "Demonstration of a Speech Processing System Consisting of a Speech Analyzer, Translator, Typer and Synthesizer," *JASA*, Vol. 34, No. 10, 1962.

126. Olson, H. F., and H. Belar, "Performance and a Code-Operated Speech Synthesizer," *JASA*, Vol. 38, No. 5, 1964.

127. Paul, A. P. House, A. S. and K. N. Stevens, "Automatic Reduction of Vowel Spectra: An Analysis-by-Synthesis Method and its Evaluation," *JASA*, Vol. 36, No. 2, 1964.

128, Peterson, G. E., "Design of Visible Speech Devices," *JASA*, Vol. 26, No. 3, 1954.

129. Peterson, E., and F. S. Cooper, "Peakpicker: A Band-Width Compression Device," *JASA*, Vol. 29, No. 6, 1957.

130. Peterson, G. E. and I. Lehiste, "Identification of Filtered Vowels," *JASA*, Vol. 31, No. 6, 1959.

131. Peterson, G. E., Sivertsen, E., and D. L. Subrahmanyam, "Intelligibility of Diphasic Speech," *JASA*, Vol. 28, No. 3, 1956.

132. Petrick, S. R., "Talking to a Computer," *New Scientist*, No. 235, 1961.

133. Phyie, D. L. and I. E. Toffier, "Some Features of the Army Channel Vocoder," *JASA*, Vol. 35, No. 12, 1963.

134. Pick, G. G.,Gray, C. B., and D. B. Brick, "The Solenoid Array - a New Computer Element," *IEEE Trans. Electronic Computers*, EC-13, No. 1, 1964.

135. Pinson, E. N., "Pitch-Synchronons Time-Domain Estimation of Formant Frequencies and Bandwidth," *JASA*, Vol. 35, No. 8, 1963.

136. Pollack, I., and L. Picket, "Effect of Noise and Filtering on Speech Intelligibility at High Levels," *JASA*, Vol. 29, No. 12, 1957.

137. Potter, R. K., Knopp, G.A., and H. C. Green, "Visible Speech," Van Nostrand, New York, 1947. /398

138. Potter, R. K. and J. C. Steinberg, "Toward the Specification of Speech," *JASA*, Vol. 22, No. 6, 1950.

139. Proceedings of the Speech Communication Seminar, Stockholm, August 29 to September 1, 1962, Publ. Speech Transmission Lab. Royal Inst. Technology, Stockholm, 1964.

140. Pruzansky, S., and P. D. Briener, "Automatic Talker Recognition Using Time-Frequency Pattern Matching," *JASA*, Vol. 33, No. 6, 1961.

141. Pun, L., "The Phonetograph," *Control*, Vol. 7, January 1963.

142. Rader, C., "Spectra of Vocoder-Channel Signals," *JASA*, Vol. 35, No. 5, 1963.

143. Rader, C. M., "Vector Pitch Detection," *JASA*, Vol. 36, No. 10, 1964.

144. Rawley, I. R., "Converting Digital Data to Volce," *Electronic Industr.* Vol. 23, No. 4, 1964.

145. Rais, A., "Vowel Recognition in Clipped Speech," *Nature*, Vol. 181, No. 3, 1958.

146. Righini, G. U., "A Pitch. Extractor of the Voice," *Acustica*, Vol. 13, No. 4, 1963.

147. Rosen, G., "Dynamic Analog Speech Synthesizer," *JASA*, Vol. 30, No. 3, 1958.

148. Sackschewsky, V. E., and H. L. Oestreicher, "Pattern Recognition as a Problem in Decision Theory and an Application to Speech Recognition," *IEEE Trans. Military Electronics*, MIL-7, No. 2-3, 1963.

149. Sakai, T., Dochita, S., Nagata, K. I., "Phonetic Typewriter," *JASA*, Vol. 35, No. 7, 1963.

150. Sakai, T., and S. I. Inone, "New Instruments and Methods for Speech Analysis," *JASA*, Vol. 32, No. 4, 1960.

151. Schief, R., "Koinzidenz-Filter als Modell fur das menschiche Tonbohenunterscheidungsvermogen," *Kybernetik*, Vol. 2, H. L, 1963.

152. Scholtz, P. N. and R. Bakis, "Spoken Digit Recognition Using Vowel-consonants Segmentation," *JASA*, Vol. 34, No. 1, 1962.

153. Schroder, M. R., "New Approach to Time Domain Analysis and Synthesis," *JASA*, Vol. 31, No. 6, 1959.

154. Schroder, M. R., and T. H. Crystal, "Auto-Correlation Vocoder," *JASA*, Vol. 32, No. 7, 1960.

155. Schroder, M. R., and E. E. David, "A Vocoder for Transmitting 10 kc/s Speech Over a 3.5 kc/s Channel," *Acustica*, Vol. 10, No. 1, 1960.

156. Seki, H., "A New Method of Speech Transmission by Frequency Demultiplication and Multiplication," *J. Acoust. Soc.* Japan, 14 June 1958.

157. Siedler, G., "Untersuchungen uber die Bedeutung bestimmter Tonfrequenzbander fur die Verstandlichkeit synthetischer Sprache und uber Anderung der Sprachverstandlichkeit bei Kanalvertauschungen," *Z. angew. Phys.* Vol. 13, Nr. 6, 1961.

158. Simmons, P. L., "Automation of Speech, Speech Synthesis and Synthetic Speech. A Bibliographic Survey from 1950-1960," *IRE Trans. Audio*, AU-9, November-December 1961.

159. Slaymaker, F. H., "Bandwidth Compression by Means of Vocoder," *IRE Trans. Audio*, AU-9, January-February 1960.

160. Smith, C. R., "A Phoneme Detector," *JASA*, Vol. 23, No. 4, 1951.

161. Smith, C. R,, "The Analysis and Automatic Recognition of Speech Sounds," *Electronic Engng*. Vol. 24, No. 8, 1962.

162. Smith, S. L., "Man-Computer Information Transfer," *Electro-Technology*, Voo. 72, August 1963.

163. Speech Recognition Gets Push From Synthesizer, *Electronics*, Vol. 34, No. 16, 1961.

164. Steele, K. W., and L. E. Cassel, "Quality Improvement in the Channel Vocoder," *JASA*, Vol. 35, No. 5, 1963.

165. Stevens, K. N., "Acoustical Analysis of Speech," *JASA*, Vol. 30, No. 7, 1958.

166. Sund, H., "A Sound Spectrometer for Speech Analysis," *Trans. Royal Inst. Technology*, Stockholm, No. 112, 1957.

167. Suzuki, I., Kadokawa, J., Nakata, K., "Formant-Frequency Extraction by the Method of Moment Calculations," *JASA*, Vol. 35, No. 9, 1963.

168. Suzuki, I., and K. Nakata, "Phonemic Classification and Recognition of Japanese Monosyllables," *J. Radio Res. Labs.*, Vol. 10, No. 49, 1963.

169. Suzuki, I., Nakata, K., and K. Maezono, "Speech Data Analyzed by Computer Program, Classification and Recognition of Japanese Monosyllables," *IEEE Trans. Military Electronics*, MIL-7, No. 2-3, 1963.

170. Tarneczy, T. H., "Vowel Formant Bandwidths and Synthetic Vowels," *JASA*, Vol. 34, No. 6, 1962.

171. Thierney, J. Gold, B. Sferrino, V., Dumanian, I. A., and E. Aho, "Channel Vocoder with Digital Pitch Extractor," *JASA*, Vol.36, No. 10, 1964.

172. Toffler, I. E., and F. B. Wade, "Pitch Extractor, Using Clippers," *JASA*, Vol. 36, No. 5, 1964.

173. Uhr, L., "Recognition of Letters, Pictures and Speech by a Discovery and Learning Program," *WESCON Techn. Papers*, Vol. 8, p. 4, August 25-28, 1964.

174. Vilbig, F., "An Analysis of Clipped Speech," *JASA*, Vol. 27, No. 1, 1955. /399

175. Vilbig, F., "Speech Compression," *JASA*, Vol. 28, No. 1, 1956.

176. Vilbig, F., "Improvement of Simplification of the Scanvocoder and its Connection to a Correlation Pulse Code System," *JASA*, Vol. 23, No. 4, 1956.

177. Vilbig, F., and K. H. Haase, "Uber einige Systeme zur Sprachbandkompression, Nachrichtentechn. Fachber." NTF, Vol. 3, 1956.

178. Vilbig, F., and K. H. Haase, "Some Systems for Speech-Band Compression," *JASA*, Vol. 28, No. 4, 1956.

179. Wallace, J. C., "Comparative Evaluation of Pitch-Signal Indicators," *JASA*, Vol. 35, No. 5, 1963.

180. Waller, R., "Self-programming Pattern Recognizer, Measurements and Control," No. 3, 1964.

181. Waller, R., "Sceptron," *J. Scient. Instruments*, Vol. 41, No. 5, 1964.

182. Weiss, M. R., and C. M. Harris, "Computer Technique for High-Speech Extraction of Speech Parameters," *JASA*, Vol. 35, No. 2, 1963.

183. Widrow, B., Groner, G. F., Hu, M. I. C., Smith, F. W., Specht, D. F., and L. R. Talbert, "Practical Applications for Adaptive Data-Processing Systems," *WESCON Techn. Papers*, Vol. 7, No. 7, 1963.

184. Winckel, F., "Tonhohenextractor fur Sprache mit Gleichstromanzeige," *Phonetica*, Nr. 3-4, 1964.

185. Wiren, I., and H. L. Stubbs, "Electronic Binary Selection System for Phoneme Classification," *JASA*, Vol. 28, No. 6, 1956.

186. Wood, D. E. and T. L. Hewitt, "New Instrumentation for Making Spectrographic Pictures of Speech," *JASA*, Vol. 35, No. 8, 1963.

187. Wood, D. E., "New Display Formant and a Flexible-Time Integration for Spectral-Analysis Instrumentation," *JASA*, Vol. 36, No. 4, 1964.

188. W. S., Dr., "Identification des personnes par la spectrographie vocale," *Automatisme*, Vol. 8, No. 7-8, 1963.

189. Yoshima, T., "Japan Firm Builds Spoken Voice Digit Recognizer," *Electronic News*, Vol. 8, No. 390, 1963.

190. Zwicker, E., "Moglichkeiten zur Spracherkennung uber den Tastsinn mit Hilfe eines Funktionsmodells des Gehors," *Elektronische Rechenaniagen*, Vol. 6, H. 6, 1964.